

# Bone Marrow Pathology Detection\*

Michael Freeman, Emmanuel Peters, Ariana Schindler, and Ziwen Zhou

*CSCI2470*

*Department of Computer Science*

*Brown University, Providence, RI, 02912*

(Dated: December 8, 2022)

**Keywords:** cell classification, CNN, RNN, pathology detection, deep learning

BRIEF ABSTRACT GOES HERE!!

## I. INTRODUCTION

### A. BACKGROUND

Introduction, background, and inspiration. What problem are we solving?

We are attempting to classify the pathology status of bone marrow slides. We have over 170k samples available to us, that come from healthy patients as well as those with various pathologies. All of us in the group have a rooting interest in science & medicine, and the member of the group who proposed the idea recognized that current bone marrow analysis is often done manually and subject to human errors. It's also known to be quite a difficult and subtle task, so we propose that a deep learning approach can either prove superior to human results or illuminate the path toward other automated approaches. As defined, this is a classification problem, but unsupervised learning approaches could be integrated (perhaps some type of feature detection and grouping to see if any interesting patterns arise).

### B. RELATED WORK

Related work that has been completed on topic or similar model.

We did not notice any meaningful work done with the dataset we found (it's quite a new dataset, which is exciting). We did draw some inspiration from work done upon a similar dataset (although it was a more restricted set that was classifying solely leukemia status). It can be found here [1]. There are no public implementations of the specific task we are trying to accomplish, and not for this dataset. However, there is an interesting technique implemented [2] for a landmark classification task that we think would be very interesting.

While doing preliminary research, we came across this article from Nature [3]. In this article, they are discussing training CNNs and RNNs on biopsied samples of the stomach and colon in order to better classify epithelial tumors. In this study, they worked to classify by tumor type as well as stage. They were able to achieve area under curves (AUCs) up to  $> 0.97$  for certain tumor types, which is very impressive. The article shows high generalization ability, and demonstrates the potential to use such a model in a diagnostic workflow system for pathologists.

Reading this article informed us that AUCs of that number were possible and might be incorporated into a reach goal. It should be noted that this study worked exclusively with diseased samples of a completely different cell type.

## II. METHODOLOGY

### A. DATA

Discussing the dataset collection and details.

As taken from the corpus description, this dataset [1] consists of 170k+ anonymized bone marrow smears from 945 patients that have been expertly annotated. Samples were collected at the Munich Leukemia Laboratory and post-processed using software from Helmholtz Munich. As it's 170k samples, we expect to do significant preprocessing on it (especially because they only arise for 945 patients, we need to ensure we're not introducing bias from sample selection).

### B. MODEL

Model choice(s), what it does and how we built it. Splitting & preprocessing, EDA, etc.

This is the least fleshed out section, as we are 1470 Capstone/2470 students trying to implement our own project. We think that starting with CNNs and RNNs first is appropriate. They were the models use in the

---

\* mfreema3; apeter13; aschind2; zzhou65

Nature project of study, and are the classic techniques we’ve studied in class for classification tasks. We are currently working on preprocessing first, but will refine our methodology summary as we attempt to create CNNs and RNNs and gauge their success. We can have Multilayer perceptrons, KNNs, and SVMs in our back pocket if our initial approaches do not prove promising.

### C. METRICS

Discuss primary metric.

As our general task is a classification one, accuracy is our primary metric. We intend to partition 10-15,000 of our samples in the dataset to treat as a testing set and another 10-15,000 samples as a validation set. AUC is a possible metric we can use, but a more familiar form of accuracy metric is also plausible.

In terms of base, target, and stretch goals, we’ve largely been considering accuracy/AUC cutoffs. For base, it would be nice to see accuracy  $> 0.7$ ; for our target, we would like to see accuracy  $> 0.85$ ; and for our stretch we would ideally like to see accuracy  $> 0.95$ . Somewhere within our target and stretch goals, we would also like to incorporate feature detection and unsupervised learning as a secondary portion of our project.

### III. RESULTS

Model interpretation

### IV. ETHICS

*Who are the major “stakeholders” in this problem, and what are the consequences of mistakes made by your algorithm?*

The major stakeholders of our bone marrow pathology project would be pathologists responsible for classifying bone marrow patients as well as their patients. Mistakes made by the algorithm would primarily be felt by the patients, who could proceed with an inaccurate diagnosis. We are much more concerned about Type II errors (false-negatives), which could delay lifesaving treatment. Type I errors (false-positives) are not ideal, but less deleterious.

*Why is Deep Learning a good approach to this problem?*

Bone marrow smear pathology detection is a very difficult task that when performed by humans relies on various heuristics. It can be error-prone, so we expect a deep learning approach to be able to learn the existing heuristics as well as see more high-level patterns that will enable more accurate classification.

### V. REFLECTION

We can answer the reflection questions in the respective subcategories.

- *How do you feel your project ultimately turned out? How did you do relative to your base/target/stretch goals?*
- *Did your model work out the way you expected it to?*
- *How did your approach change over time? What kind of pivots did you make, if any?*
- *Would you have done differently if you could do your project over again?*
- *What do you think you can further improve on if you had more time?*
- *What are your biggest takeaways from this project/what did you learn?*

#### A. CHALLENGES

Challenges we ran into, limitations of implementations.

#### B. ACCOMPLISHMENTS

Accomplishments we are proud of.

### VI. WHAT WE LEARNED

Recap

### VII. FURTHER RESEARCH OPPORTUNITIES

### VIII. ACKNOWLEDGMENTS

Possible section to thank for inspiration, guidance, technical support, mentorship, or contributions over the duration of the research.

- 
- [1] Bone Marrow image dataset used for analysis  
<https://www.kaggle.com/datasets/andrewmvd/bone-marrow-cell-classification>
- [2] Landmark classification task implemented  
<https://github.com/cvdfoundation/google-landmark>
- [3] Deep Learning Models for Histopathological Classification of Gastric and Colonic Epithelial Tumours  
<https://www.nature.com/articles/s41598-020-58467-9>